PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 5/06, 9/00

A1

(11) International Publication Number: **WO 97/33273**

(43) International Publication Date: 12 September 1997 (12.09.97)

(21) International Application Number: PCT/US97/04350

(22) International Filing Date: 6 March 1997 (06.03.97)

(30) Priority Data:
08/613,269 8 March 1996 (08.03.96) US

(71) Applicant: MOTOROLA INC. [US/US]; 1303 East Algonquin Road, Schaumburg, IL 60196 (US).

(72) Inventor: FINEBERG, Adam, B.; 12761 Saratoga Glen Court, Saratoga, CA 95070 (US).

(74) Agents: WOOD, J., Ray et al.; Motorola Inc., Intellectual Property Dept., 1303 East Algonquin Road, Schaumburg, IL 60196 (US).

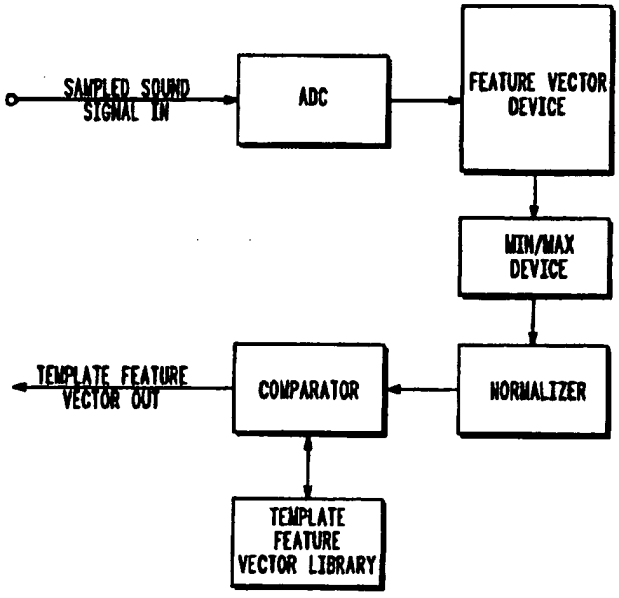
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD AND RECOGNIZER FOR RECOGNIZING A SAMPLED SOUND SIGNAL IN NOISE

(57) Abstract

A sound recognizer uses a feature value normalization process to substantially increase the accuracy of recognizing acoustic signals in noise. The sound recognizer includes a feature vector device (110) which determines a number of feature values for a number of analysis frames, a min/max device (120) which determines a minimum and maximum feature value for each of a number of frequency bands, a normalizer (130) which normalizes each of the feature values with the minimum and maximum feature values resulting in normalized feature vectors, and a comparator (140) which compares the normalized feature vectors with template feature vectors to identify one of the template feature vectors that most resembles the normalized feature vectors.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

METHOD AND RECOGNIZER FOR RECOGNIZING A SAMPLED SOUND SIGNAL IN NOISE

Field of the Invention

The present invention relates, in general, to sound recognition, and in particular, to sound recognition in a high or variable noise environment.

Background of the Invention

Advancing technology is moving steadily towards commercializing sound recognition by electronic devices such as speech recognizers. Generally there are two types of speech recognizers. One performs certain operations when the user gives short commands. A second type accepts dictated speech and enters the speech as text.

Most speech recognizers must be trained by the user before they can recognize words or phrases spoken by the user. These are termed "speaker dependent" speech recognizers, meaning that the speech recognizer must be trained by the user's voice before the recognizer can interpret user words and commands. Training a speech recognizer requires a user to speak certain words or phrases into the recognizer, usually many times, so that the speech recognizer can recognize the user's speech pattern. Later when the user is using the speech recognizer, the speech recognizer will compare the input voice signal with various stored speech templates to find a template that most resembles the input voice signal. This method is called "pattern matching".

A user will generally "train" a speech recognizer in an environment that has relatively low interfering noise. Subsequently, most speech recognizers must be used in environments of low interfering noise. Otherwise, the speech

recognizer will not be able to separate spoken words from background noise. Where speech recognizers are used in low noise environments, a fairly high rate of recognition is achieved. If the speech recognizer is trained in a location having a moderate, constant background noise, and subsequently used in an environment that has the same moderate, constant background noise, a high recognition rate is achieved. However, when these speech recognizers are used in high noise environments with negative signal-to-noise ratios and environments where the noise present is different than the background noise present in the training session, the recognition rate falls to very low, unusable accuracy levels.

To correct the problem of background noise, conventional speech recognizers attempt to estimate the characteristics of the surrounding noise and then determine the effects on the user's voice. Various techniques are incorporated to build statistical or parametric models of the noise which are subtracted from the sound signal. In high and variable noise environments, these models are very inaccurate.

Brief Description of the Drawings

FIG. 1 is a block diagram of a voice recognizer according to a preferred embodiment of the present invention.

FIG. 2 shows a flow diagram of a preferred embodiment of the present invention.

FIG. 3 shows a flow diagram of a method used to calculate feature values according to a preferred embodiment of the present invention.

FIG. 4 is a representation of a power spectrum of a sampled sound signal with frequency filters imposed thereon according to the present invention.

FIG. 5 shows a matrix of features of the sampled sound signal according to a preferred embodiment of the present invention.

FIG. 6 shows a matrix of normalized features for the features of FIG. 3 according to a preferred embodiment of the present invention.

Detailed Description of the Invention

A preferred embodiment of the present invention is used in robust voice recognition for sound recognizers. The preferred embodiment is well suited for use in cellular phones in automobiles where a user can keep both hands on the steering wheel, eyes on the road, and still make a phone call even with windows down and the stereo system on loud. Unlike conventional speech recognizers having an unusable, poor accuracy rate in high and/or variable noise conditions, sound recognizers designed according to the preferred embodiment of the present invention are robust and can obtain a very high accuracy rate in environments which have variable noise and noise levels greater than the volume of the user's speech.

The present invention shall be described hereafter in conjunction with the drawings. In particular, the preferred embodiment shall be described with reference to FIG. 1 in combination with the other figures.

The present invention may be applied to the recognition of any acoustic sound. For instance, the acoustic sound may be speech, grunting sounds, sounds made by animals, sounds made by instruments including percussion instruments, or any other type of sound. Most commonly, the present invention relates to recognition of speech.

FIG. 1 shows a sound recognizer 100 according to a preferred embodiment of the present invention. In the preferred embodiment, an acoustic signal is input to an analog to digital converter (ADC) 105 of sound recognizer 100 where the signal is converted to a digital signal and sampled at a rate

of 16KHz. Other sampling rates may be used as appropriate, such as 8KHz.

The sampled digital signals are input to feature vector device 110 which divides the sampled digital signals into analysis frames. Each analysis frame can be chosen to be either of fixed time width (such as 20ms) or may be of varied time width depending upon signal characteristics such as pitch periods or other determining factors. The starting point of each analysis frame can be chosen to be either before, at, or after the end point of the previous frame. In the preferred embodiment, the analysis frames are chosen to be of fixed time width, and each analysis frame starts at the ending point of the previous analysis frame.

For each of the analysis frames, feature vector device 110 computes a feature vector (210 of the flow chart of FIG. 2). For any given number of analysis frames, feature vector device 110 generates an equal number of feature vectors. A feature vector is a series of values, or plurality of feature values, which are derived from the sampled sound signal within a given analysis frame. These feature values are representative of the information contained in the sampled sound signal.

There are many techniques known to those skilled in the art of speech recognition which may be used to determine feature vectors. The techniques include Linear Predictive Coding (LPC) Coefficients, Cepstral Coefficients, Log Area Ratios, and Mel Scale Filterbank Coefficients. The preferred embodiment of the present invention utilizes the Mel Scale Filterbank Coefficients method, although the present invention will operate with other feature vector techniques, such as those listed above.

Mel Scale Filterbank Coefficients are computed in the following manner with reference to the flow chart of FIG. 3.

1. The sound signal samples for an analysis frame are passed through a high frequency pre-emphasizing filter to

whiten the spectrum of the sound signal samples (310 of the flow chart of FIG. 3). This increases the relative energy in the high frequency components as compared to the energy of the low frequency components. Benefits are obtained when the preferred embodiment of the present invention is used with speech signals since low frequency components of speech have a relative energy greater than high frequency components and the two components are re-balanced in the pre-emphasizing filter. In the preferred embodiment, the filtering is accomplished according to the equation:

$$p_i(k) = s_i(k) - s_i(k-1)$$

where $s_i(k)$ is the sound signal sample at position k in analysis frame "i", $s_i(k-1)$ is the sound signal sample in analysis frame "i" at the previous position in time "k-1", and $p_i(k)$ is the pre-emphasized sound signal sample at position "k" in analysis frame "i". One skilled in the art of speech recognition will recognize that other pre-emphasis filters may be used.

2. The pre-emphasized sound signal samples for each analysis frame are bandpass filtered by a series of filters covering different frequency bands. The filters may be applied in any computational manner desired in either the time domain or the frequency domain. In the preferred embodiment, the filters are applied in the frequency domain. First, however, a power spectrum of the pre-emphasized sound signal samples in the analysis frames must be computed (320 of FIG. 3). The power spectrum is found by:

a. The pre-emphasized sound signal samples in the analysis frame are multiplied by samples of a window function, or weighting function. Any window function may be applied. For purposes of explaining the present invention, a simple rectangular window is assumed (the window has a value of 1.0 for all samples).

b. The Fourier Transform of the pre-emphasized sound signal samples in each windowed analysis frame is computed.

c. Values for the power spectrum are obtained by squaring the Fourier Transform values.

After the values for the power spectrum are determined, the band-pass filters are applied in the frequency domain by a filter weighting value for each of the power spectrum values (330 of FIG. 3). Although many filter weighting functions may be used within the band-pass filters, the preferred embodiment incorporates a raised cosine weighting profile which can be seen in FIG. 4.

FIG. 4 shows a power spectrum 400 having the raised cosine profile 410 imposed thereon. The frequency bands for each band-pass filter, or raised cosine profile 410, in the preferred embodiment of the present invention are laid out along the frequency axis according to a Mel or Bark scale which approximates the frequency response of the human ear. The frequency bands for the band-pass filters (raised cosine profiles 410) are approximately linearly spaced from 0 to 1KHz, and logrithmically spaced above 1KHz. Filter spacings other than those defined for the preferred embodiment may also be used. As seen in FIG. 4 for the preferred embodiment, the band-pass filters, or raised cosine profiles 410, overlap. The outputs for the band-pass filters, or raised cosine profiles 410, are computed according to:

$$f_{ij} = \Sigma P_i(\omega)B_j(\omega)$$

where $P_i(\omega)$ is the powe spectrum value for analysis frame "i" at frequency ω, $B_j(\omega)$ is the band-pass filter weighting function, or frequency response for filter "j" at frequency ω, Σ represents the summation operation over all frequencies ω, and f_{ij} is the band-pass filter output for analysis frame "i" and band-pass filter "j".

After all the band-pass filter outputs for each analysis frame "i" ($0 \leq i \leq n$) and each band-pass filter "j" ($0 \leq j \leq m$) have been calculated, feature vector device 110 calculates feature values "v_{ij}" of the sampled sound signal by taking the log of each band-pass filter f_{ij} (340 of FIG. 3). The result may be shown as a matrix such as the one illustrated in FIG. 5 structured with "i" analysis frames and "j" band-pass filters having a dimension of n x m. All of the feature values within an analysis frame, v_{i1} through v_{im}, form a single feature vector (as in v_{11} through v_{1j}, item 510), and all analysis frames, $0 \leq i \leq n$, form a plurality of feature vectors for the sampled sound signal.

Once the plurality of feature vectors for analysis frames "i" = 1 through "n" have been computed, a min/max device 120 (FIG. 1) coupled to or incorporated within feature vector device 110 reviews all of the feature values within a frequency band for band-pass filter "j" and finds the minimum (min_j) feature value and the maximum (max_j) feature value for frequency band "j" for all analysis frames, $0 \leq i \leq n$ (220 of FIG. 2). These minimum and maximum values are used to determine normalized feature values, "v~".

Normalizer 130 of FIG. 1 is coupled to min/max device 120 and feature vector device 110. Normalizer 130 normalizes each of the feature values across a frequency band, or band-pass filter, "j", with the minimum and maximum feature values for that band-pass filter to determine the normalized feature values "v~" (230 of FIG. 2). The normalization equation is:

$$v\sim_{ij} = (v_{ij} - min_j)/ (max_j - min_j)$$

where $v\sim_{ij}$ is one of the normalized feature values, v_{ij} is one of the feature values, min_j is the minimum feature value for the "j^{th}" frequency band, and max_j is the maximum feature value for the "j^{th}" frequency band.

In an improved alternative embodiment, the min/max device 120 finds a weighted minimum (min_j) feature value and a weighted maximum (max_j) feature value for frequency band

"j" for all analysis frames, $0 \leq i \leq n$ (220 of FIG. 2). These minimum and maximum values are computed as follows:

$$\omega_{minj} = \sum_{k=-r}^{r} weight[K]min[j+k]$$

$$\omega_{maxj} = \sum_{k=-r}^{r} weight[K]max[j+k]$$

where r is ususally a small value such as 2 or 3, weight[k] is typically a weighting function where the central point, weight[0], has a value greater than or equal to all other weight values and

$$\sum_{k=-r}^{r} weight[K] = 1.0.$$

These weighted minimum and weighted maximum values are used to determine the normalized feature values, "v~". In this embodiment the normalization equation is:

$$v\sim_{ij} = (v_{ij} - \omega_{minj})/ (\omega_{maxj} - \omega_{minj})$$

where, ω_{minj} is the weighted minimum feature value for the "j^{th}" frequency band, and ω_{maxj} is the weighted maximum feature value for the "j^{th}" frequency band.

The result of the normalization process can be shown as a matrix as illustrated in FIG. 6. Each of the analysis frames "i" of FIG. 6 represents a normalized feature vector (610).

Comparator 140 of FIG. 1 is coupled to normalizer 130 and compares the normalized feature vectors with template feature vectors to determine which of the template feature vectors most resembles the normalized feature vectors. Sets of template feature vectors representing phrases or commands are stored within template feature vector library 150. Comparator 140 compares the normalized feature vectors from normalizer 130 with each of the template feature vectors in template feature vector library 150 in turn (240 of FIG. 2) and determines which set of the template feature vectors most resembles the normalized feature vectors (250). This is done

by computing a distance metric between the normalized feature vectors and each of the sets of template feature vectors. The set of template feature vectors having the minimum distance metric is determined to be the one which most resembles the normalized feature vectors. Comparator 140 of FIG. 1 outputs as a best-fit match the set of template feature vectors from template feature vector library 150 which most resembles (has the minimum distance metric) the normalized feature vectors (250 of FIG. 2).

There are several well known methods whereby the plurality of normalized feature vectors may be compared with the template feature vectors to find a best-fit match. Studies using the preferred embodiment of the present invention show that comparing the plurality of normalized feature vectors with the template feature vectors in a dynamic time warping process yields the best results.

As mentioned earlier, the present invention when used in a speaker dependent, small vocabulary sound recognition system is very robust and increases recognition accuracy in high and variable noise environments from unusable accuracy rates to very high accuracy rates.

It should be recognized that the present invention may be used in many different sound recognition systems. All such varied uses are contemplated by the present invention.

What is claimed is:

CLAIMS:

1. A method comprising the steps of:

computing feature values for a number of analysis frames of an acoustic signal, the feature values computed over a plurality of frequency bands;

for each of the plurality of frequency bands, determining which of the feature values within a respective one of the plurality of frequency bands is a minimum feature value and which of the feature values within the respective one of the plurality of frequency bands is a maximum feature value;

comparing each of the feature values within each of the plurality of frequency bands with the minimum feature value and the maximum feature value of the respective one of the plurality of frequency bands to obtain normalized feature values, wherein all of the normalized feature values for a given one of the number of analysis frames defines one of a plurality of normalized feature vectors;

comparing the plurality of normalized feature vectors with sets of template feature vectors to determine template feature vectors that most resemble the plurality of normalized feature vectors; and

outputting the template feature vectors that most resemble the plurality of normalized feature vectors.

2. A method according to claim 1 wherein the step of determining which of the feature values within a respective one of the plurality of frequency bands is a minimum feature value and which of the feature values within the respective one of the plurality of frequency bands is a maximum feature value comprises finding a weighted minimum feature value and a weighted maximum feature value for each of the plurality of frequency bands.

3. A method according to claim 1 wherein the step of computing feature values comprises:

dividing a power spectrum of the acoustic sound signal into the plurality of frequency bands;

weighting the power spectrum within each of the plurality of frequency bands according to a weighting function to obtain filter outputs; and

calculating the feature values from the filter outputs.

4. A method according to claim 3 wherein the power spectrum is weighted in a frequency domain according to a raised cosine profile.

5. A method according to claim 4 wherein the raised cosine profile calculates the filter outputs according to an equation:

$$f_{ij} = \Sigma P_i(\omega)B_j(\omega)$$

where f_{ij} is one of the filter outputs at an analysis frame "i" of the number of analysis frames and a frequency band "j" of the plurality of frequency bands, $P_i(\omega)$ is a power spectrum value for the analysis frame "i" at frequency ω, $B_j(\omega)$ is a band-pass filter weighting function, or frequency response, for the frequency band "j" at frequency ω, and Σ represents a summation operation over all frequencies ω.

6. A method according to claim 1 wherein the step of comparing the normalized feature vectors with sets of template feature vectors to determine the template feature vectors that most resemble the normalized feature vectors includes comparing the normalized feature vectors with the sets of template feature vectors in a dynamic time warping process.

7. A method according to claim 3 wherein the step of calculating the feature values from the filter outputs comprises taking a log of each of the filter outputs.

8. A method according to claim 3 wherein the power spectrum is divided into the plurality of frequency bands according to a Mel or Bark scale to approximate a frequency response of a human ear.

9. A method according to claim 1 wherein each of the normalized feature vectors is found according to an equation:

$$v\sim_{ij} = (v_{ij} - \min_j)/ (\max_j - \min_j)$$

where $v\sim_{ij}$ is one of the normalized feature values of an analysis frame "i" of the number of analysis frames and a frequency bands at "j" of the plurality of frequency bands;

v_{ij} is one of the feature values at the analysis frame "i" and the frequency band "j";

$\min_j$ is the minimum feature value for the "j" frequency band; and

$\max_j$ is the maximum feature value for the "j" frequency band.

10. A sound recognizer comprising:

a feature vector device which computes feature values for a number of analysis frames of an acoustic signal input to the sound recognizer, the feature values computed for a plurality of frequency bands;

a min/max device coupled to the feature vector device to determine which of the feature values within a respective one of the plurality of frequency bands is a minimum feature value and which of the feature values within the respective one of the plurality of frequency bands is a maximum feature value for each of the plurality of frequency bands across all the number of analysis frames;

a normalizer coupled to the min/max device which compares each of the feature values within each of the plurality of frequency bands with the minimum feature value and the maximum feature value of the respective one of the plurality of frequency bands to obtain normalized feature values, wherein all of the normalized feature values for a given one of the analysis frames defines one of a plurality of normalized feature vectors; and

a comparator coupled to the normalizer which compares the plurality of normalized feature vectors with sets of template feature vectors to determine template feature vectors that most resemble the plurality of normalized feature vectors, the comparator outputting the template feature vectors that most resemble the plurality of normalized feature vectors.

11. A sound recognizer according to claim 10 wherein the normalizer computes each of the plurality of normalized feature vectors according to the equation:

$$v\sim_{ij} = (v_{ij} - min_j)/ (max_j - min_j)$$

where $v\sim_{ij}$ is one of the normalized feature values of an analysis frame "i" of the number of analysis frames and a frequency bands at "j" of the plurality of frequency bands;

v_{ij} is one of the feature values at the analysis frame "i" and the frequency band "j";

min_j is the minimum feature value for the "j" frequency band; and

max_j is the maximum feature value for the "j" frequency band.

12. A sound recognizer according to claim 10 wherein the comparator is coupled to a template feature vector library, the template feature vector library containing the sets of template feature vectors.

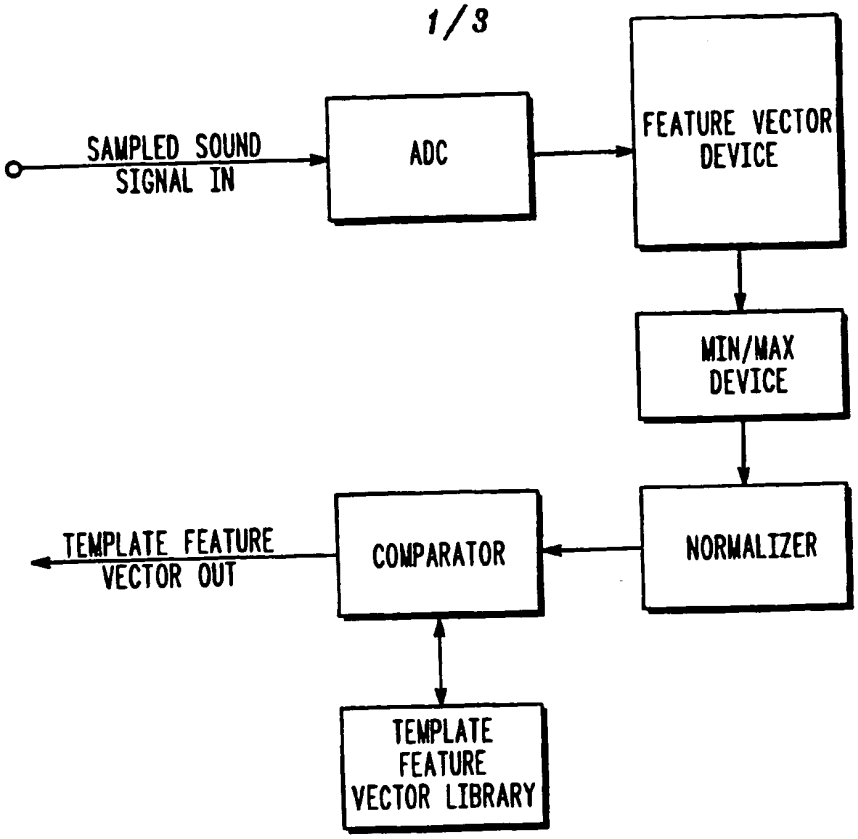

SAMPLED SOUND SIGNAL IN
ADC
FEATURE VECTOR DEVICE
MIN/MAX DEVICE
NORMALIZER
COMPARATOR
TEMPLATE FEATURE VECTOR OUT
TEMPLATE FEATURE VECTOR LIBRARY

FIG.1

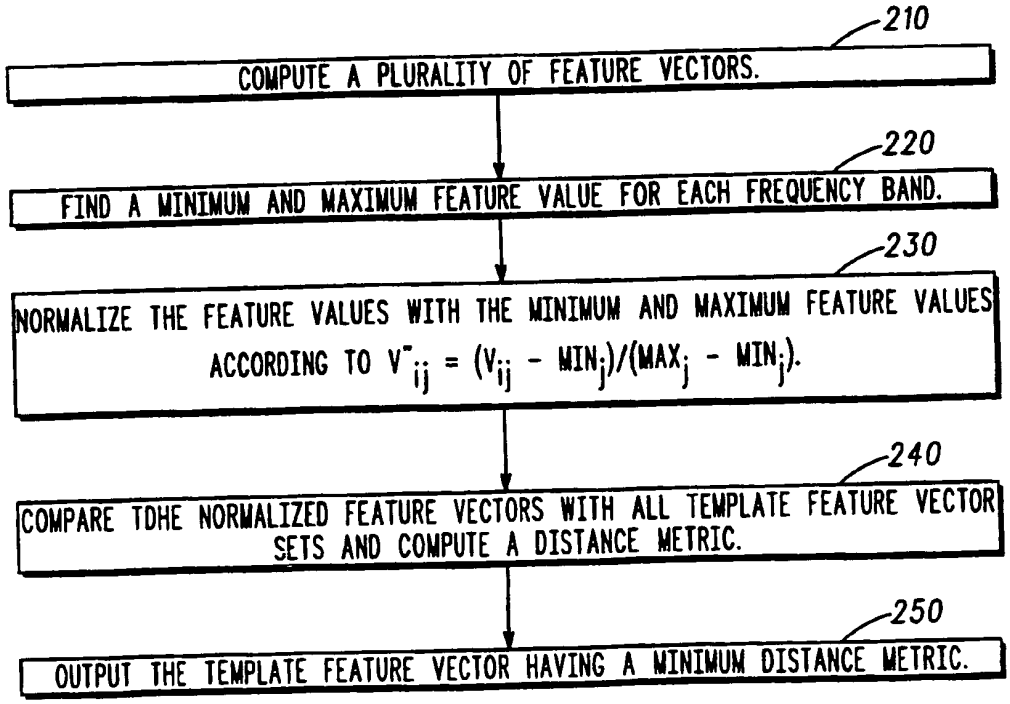

210
COMPUTE A PLURALITY OF FEATURE VECTORS.
220
FIND A MINIMUM AND MAXIMUM FEATURE VALUE FOR EACH FREQUENCY BAND.
230
NORMALIZE THE FEATURE VALUES WITH THE MINIMUM AND MAXIMUM FEATURE VALUES ACCORDING TO $V^{-}_{ij} = (V_{ij} - MIN_j)/(MAX_j - MIN_j)$.
240
COMPARE TDHE NORMALIZED FEATURE VECTORS WITH ALL TEMPLATE FEATURE VECTOR SETS AND COMPUTE A DISTANCE METRIC.
250
OUTPUT THE TEMPLATE FEATURE VECTOR HAVING A MINIMUM DISTANCE METRIC.

FIG.2

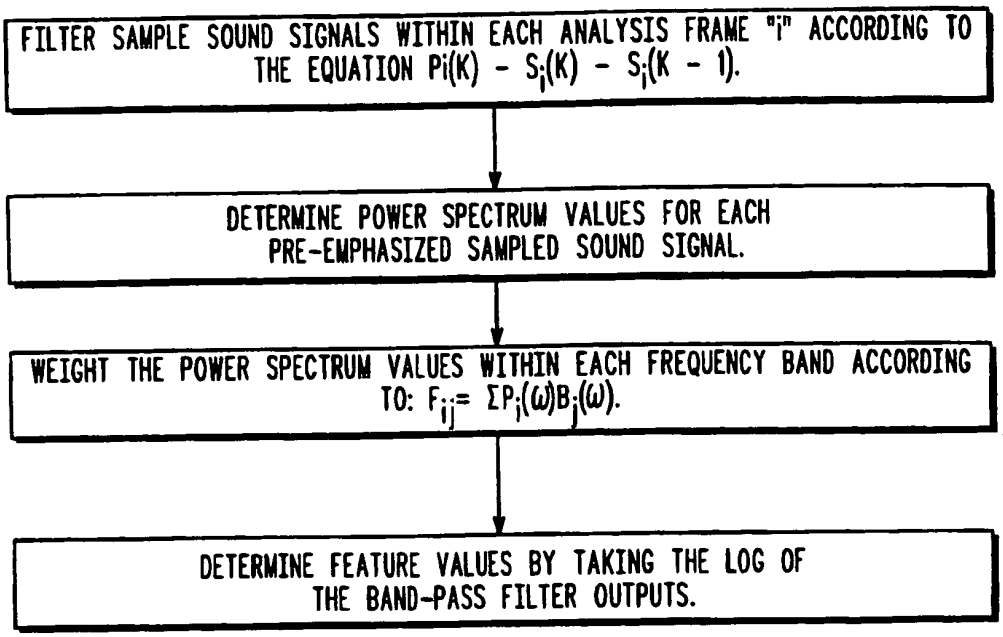


FIG.3

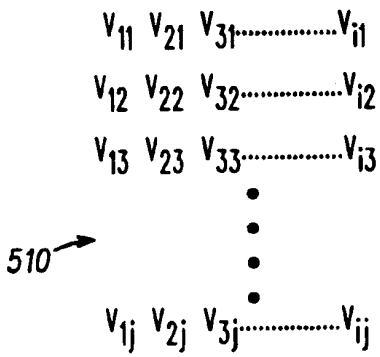


FIG.5

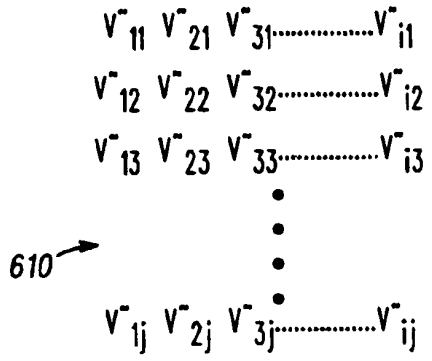


FIG.6

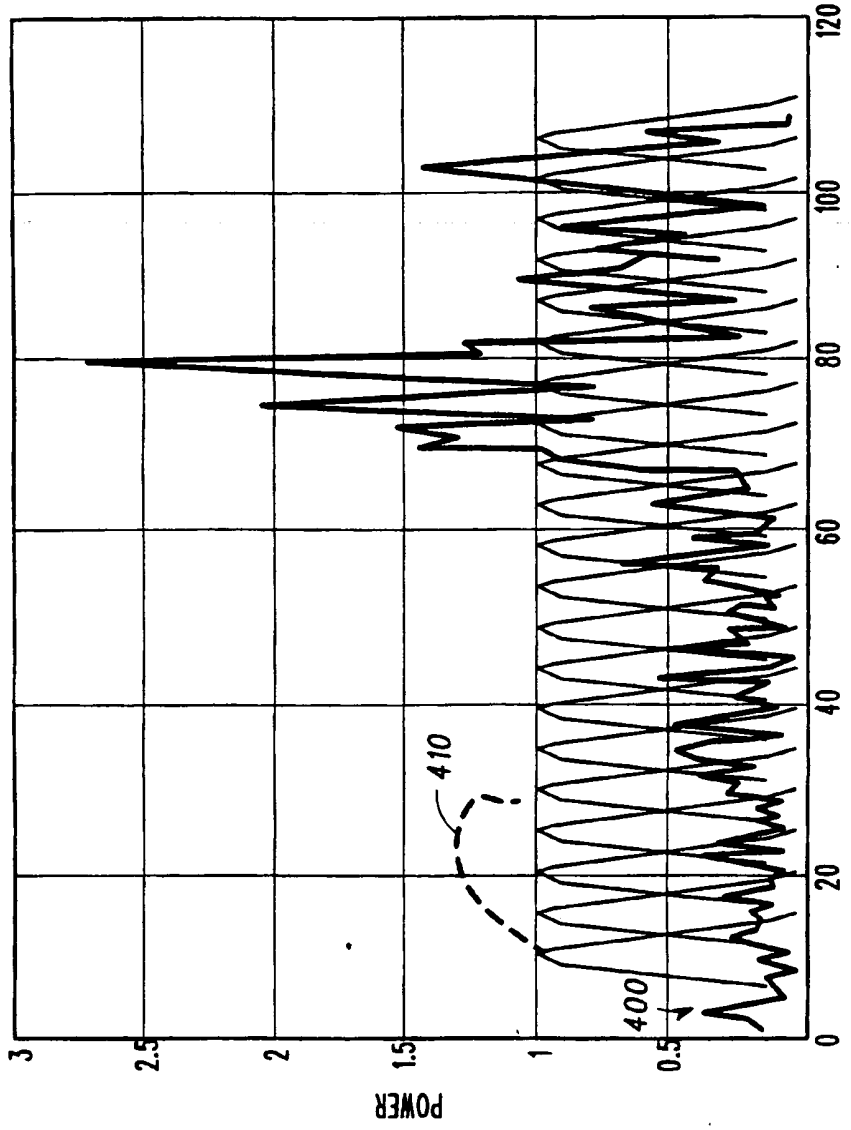


FIG. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/04350

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :395/2.42, 2.43

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/2.42, 2.43, 2.33, 2.34, 2.35, 2.36, 2.39

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, SMARTPATENT, IEEE AND IEE CDROM

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US 5,586,215 A (STORK et al) 17 December 1996, col. 6, lines 44-52, Fig. 3, col. 6 lines27-36.	1, 3, 7-8, 10
Y	US 5,023,910 A (THOMSON) 11 June 1991	2, 4, 5
Y	TSOUKALAS, D. et al, Speech Enhancement using Psychoacoustic Criteria, ICASSP '93, April 1993, Section 2, Sections 2.1-2.3, Fig.s 1-4.	1, 3, 7-8, 10
Y	HIRSCH H. G., EHRLICHER, C. Noise Estimation Techniques for Robust Speech Recognition, ICASSP '95, May 1995, pages 153-156, especially 155, equation (3).	9, 11

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
27 MAY 1997

Date of mailing of the international search report
17 JUL 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
ALLEN MACDONALD
Telephone No. (703) 305-9708

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/04350

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PARSONS, T. Voice and Speech Processing, McGraw Hill, 1987, pages 297-303 and pages 170-171 especially page 171 line 4, and pages 297 -303 especially page 29 lines 26-37.	6, 12
A	MILNER B. P., VASEGHI, S. V. Comparison of some Noise-Compensation Methods for Speech Recognition in Adverse Environments IEE Proceedings Vision, Image, and Signal Processing, October 1994, Vol 141 No. 5.	1-12
A	US 5,581,654 A (TSUTSUI) 3 December 1996	1-12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/04350

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

G10L 5/06, 9/00